

Alicia Sellés Carot
MASmedios.com
alicia@masmedios.com

Enrique Orduña-Malea
Universidad Politécnica de Valencia.
Instituto de Diseño y Fabricación (IDF-UPV).
enorma@upv.es

Jorge Serrano-Cobos
MASmedios.com
jorge@masmedios.com

Estrategias y oportunidades tecnológica

en la generación de *Linked Data* en las bibliotecas

La realidad de la gran mayoría de bibliotecas incluye cada vez más aplicaciones, servicios y tecnologías, con mayor o menor grado de integración o reutilización de datos y esfuerzos. Esta diversidad se presenta a menudo como un obstáculo en cuanto a explotación de recursos y propuestas de incorporación de nuevos estándares, lo que implica la necesidad de aplicar diversas tecnologías de integración.

1. ¿Qué es mi biblioteca?: una diversidad de plataformas y servicios

La actividad y servicio de biblioteca hace mucho tiempo que no se limita ni a un espacio físico ni a una única herramienta o plataforma web. Portales bibliotecarios y OPACs han convivido durante años sin ningún tipo de integración (como mucho un enlace entre un sitio y otro) y, a pesar de que siguen existiendo muchos casos de servicios totalmente inconexos, en los últimos años nos hemos encontrado con iniciativas que buscan una mejor atención al usuario, un servicio centralizado que vaya derivando a las plataformas correspondientes.

Los gestores de contenidos (CMSs) y otras aplicaciones han ido facilitando la gestión de información, no únicamente la bibliográfica, sino de documentos en formatos muy variados, y las bibliotecas han ampliado los canales para la prestación de servicios (Sellés-Carot, y Serrano-Cobos, 2011).

Con la aparición de la Web social, los repositorios de acceso abierto, las herramientas o servicios de control de publicaciones electrónicas, el escenario se ha complicado no sólo por el número de herramientas y plataformas a gestionar, sino también por los diferentes formatos, arquitecturas informáticas y entornos tecnológicos, esquemas de metadatos con los que funcionan, etc., cuestiones que dificultan la integración o la construcción de un servicio coherente. Por ejemplo, en un escenario plausible, podemos observar cómo una biblioteca organiza un club de lectura “físico” y lo publica en Facebook.

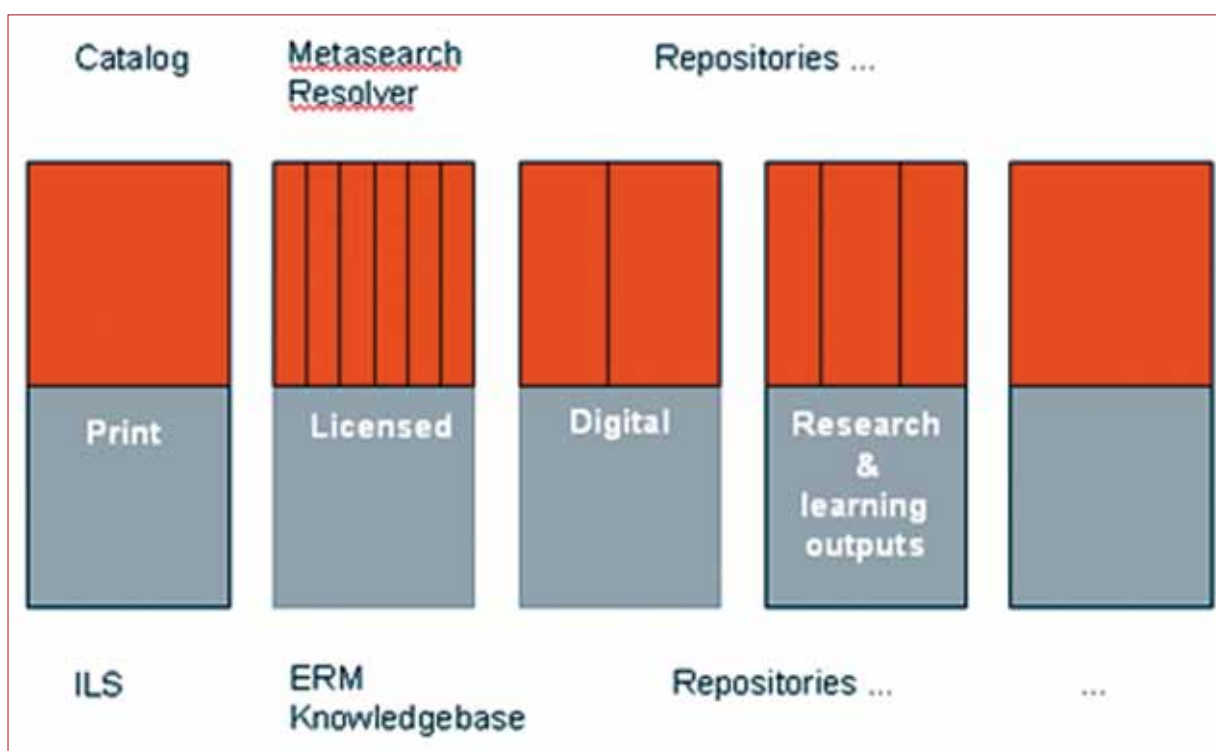
Allí confirman su asistencia “amigos” o “fans” que, como no son usuarios de la biblioteca, no sólo no pueden hacer una reserva de un ejemplar, sino que tampoco pueden participar en club, por política interna.

Con tanta diversidad de entornos y plataformas, se encuentran problemas relativos a la explotación de recursos o de generación de datos y canales eficientes de comunicación pero, sobre todo, a la formulación de propuestas para la incorporación de nuevos estándares que abran un abanico de posibilidades, que no sólo explote el trabajo realizado sino que sea capaz de generar mucho más valor al usuario final.

2. Integración: ¿por qué?, ¿de qué?, ¿cómo?

En la amplia bibliografía existente sobre integración e interoperabilidad se encuentran definiciones más o menos extensas y exhaustivas, que no son objeto de este trabajo, pero de las que se pueden extraer las cuestiones clave relativas a la integración (Dempsey, 2007):

- Interoperabilidad: los sistemas deben ser capaces de entenderse e interactuar.
- Intercambio: el resultado de la interacción de los sistemas son datos/información.
- Producción: se habla de integración cuando gestionamos o “producimos” información nueva derivada de los datos obtenidos porque los sistemas han interactuado.



Entorno del sistema bibliotecario (fuente: Lorcan Dempsey Webblog)

Es complicado hablar de integración total, pero las bibliotecas han seguido distintos caminos para integrar funcionalidades y servicios en una única plataforma, siguiendo diversas estrategias.

2.1. Estrategias de integración

Es complicado hablar de integración total, pero las bibliotecas han seguido distintos caminos para integrar funcionalidades y servicios en una única plataforma, siguiendo principalmente una de estas tres estrategias:

a) Inclusión del OPAC/SIGB y otras herramientas dentro del portal bibliotecario

El OPAC se “integra” dentro de un portal de la biblioteca, realizado con otra herramienta de CMS. Dependiendo de las posibilidades y el presupuesto, se pueden ver desde formularios de consulta o la desaparición total de la interfaz propia del OPAC original, hasta un simple enlace desde el portal al OPAC, que funciona de forma independiente.

Pero no se puede hablar de integración. Por ejemplo, en una hipotética sección de “agenda de eventos”, seguiríamos sin poder automatizar de forma realmente integrada funcionalidades como:

- La posibilidad de reservar los ítems que hay en la biblioteca de un autor al que se va a invitar la semana que viene.
- En la ficha de un registro de ese autor, incluir la información del evento.
- Que “x” usuarios de la biblioteca, que además participan en el foro online, han declarado públicamente ser fans suyos y han creado un club de lectura sobre él.

b) Portal bibliotecario dentro del SIGB o SIGB extendido.

Algunos sistemas de gestión de contenidos incluyen funcionalidades para la gestión de contenidos como noticias, eventos, boletines, etc., con lo que las funcionalidades básicas del portal quedan cubiertas en el módulo OPAC. En este caso, la integración tampoco es completa, puesto que el enfoque del SIGB sigue siendo básicamente la gestión física de la colección y necesitaría mejorar la explotación de los datos integrados en una única herramienta, que en este caso tendría que dejar de llamarse SIGB y que algunos autores han denominado SIGB exten-

dido) que sea capaz de gestionar gran parte del servicio (físico, digital, profesional...).

c) Sistema híbrido

En la incorporación al abanico de herramientas encontramos otras como metabuscadores o las nuevas herramientas que se han dado en llamar “de descubrimiento” (Next Generation Discovery Tools) (Yang y Wagner, 2010), orientadas a la recuperación y descubrimiento de información por parte del usuario pero que no integran los datos, sólo facilitan el acceso a ellos.

2.2. Principales problemas y/o retos

El reto más interesante en cuanto a opciones para explotar y generar datos de esa integración es la capacidad de trabajarlos y transformarlos. Si nuestro sistema es capaz de realizar un tratamiento de los datos que se integran, esto facilitará el trabajo para su transformación y explotación en otros ámbitos o para realizar presentaciones distintas a las que estamos habituados.

Otro retos incluyen trabajar con multitud de protocolos y estándares de integración, sistemas propietarios versus open source, protocolos de intercambio y formatos de información, o la legislación, especialmente en materia de protección de datos y de derechos de autor.

3. Oportunidades en forma de Linked Data

El tratamiento puede ser tan sencillo como modificar los datos buscando la normalización (por ejemplo, normalizar los tipos de ítems que usa cada sistema, detectar ítems/contenidos duplicados, etc.) o más complejo como detectar términos equivalentes, autoridades duplicadas, etc. La transformación, nos facilita el camino hacia estándares más abiertos como RDF.

Generar y gestionar datos en formato RDF permite a los servicios de información sacar provecho tanto consumiendo como produciendo *Linked Data* a través de la interconexión y explotación de datos que antes no estaban interenlazados, usando las

Si nuestro sistema es capaz de realizar un tratamiento de los datos que se integran, esto facilitará el trabajo para su transformación y explotación en otros ámbitos o para realizar presentaciones distintas a las que estamos habituados.

posibilidades de la Web para eliminar barreras a los usuarios para disfrutar de las infinitas potencialidades del hipertexto¹.

Además, la interconexión de datos permite generar información nueva al contextualizar datos: simplemente por su mera relación intrínseca o al generar relaciones de alto valor añadido mediante la asignación coordinada de relaciones semánticas entre los metadatos que describen a los objetos digitales, más allá de las relaciones clásicas jerárquicas y asociativas (García García, 2012).

Las ontologías, el próximo caballo de batalla que llevaría a los datos enlazados (*Linked Data*) hacia la Web semántica, necesitarán primero de datos interconectados junto a sus descripciones semánticas, lo que se puede hacer ya, sin esperar a acordar un vocabulario controlado internacional común, mediante la integración de recursos relacionados, tanto de información como de cualquier otro tipo (Bizer, Cyganiak y Heath, 2007; Pastor-Sánchez, 2013).

La publicación, utilización y reutilización de datos enlazados favorecerá la interconexión y explotación de datos anteriormente dispersos, lo que puede ser aprovechado para la elaboración y diseño de nuevos servicios de información de nueva generación. Todo ello es posible gracias a la utilización de *Linked Data*.

3.1. *Linked Data*

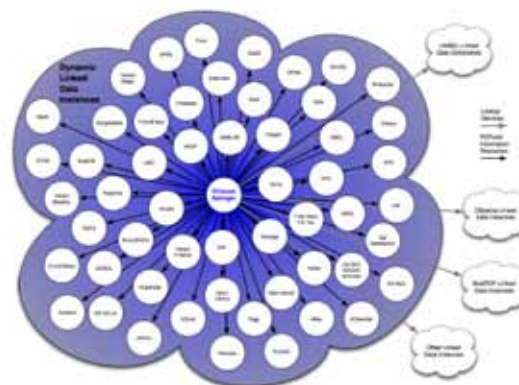
Linked Data es simplemente una guía o serie de pautas para la publicación de recursos en la Web provenientes de diferentes fuentes, de forma que las relaciones entre estos recursos queden explicitadas y, por tanto, legibles por máquina.

Las principales pautas para esta publicación “adecuada” de recursos son proporcionadas por Tim Berners-Lee²:

- Usar URIs como nombre de las cosas o recursos.
- Aprovechar el protocolo HTTP incluido en la URI, para que los usuarios puedan localizar y consultar estos recursos (a este proceso se le conoce como desreferenciar).

- Una vez que la URI se desreferencia, proporcionar información útil acerca del recurso, es decir, describirlo.
- Incluir enlaces a otras URIs, que guarden algún tipo de relación con la URI consultada, de forma que el usuario pueda navegar a través de recursos relacionados.

La idea parte de la posibilidad que dan las URIs (*Universal Resource Identifiers*) para la identificación tanto de direcciones (URLs) como de nombres (URNs). Una URI puede ser usada para identificar un recurso concreto localizado en una determinada dirección, y también para identificar un nombre o un concepto. La estructura URI permite la identificación universal de cualquier tipo de objeto.



Si esta URI se asocia con el protocolo HTTP, los usuarios podrán, a través de su navegador, acceder desde esta URI a un conjunto de recursos relacionados con ésta, conformando lo que se denomina un documento digital compuesto. Por ejemplo, una URI podría identificar a una revista y, a partir de ésta, dirigir a un artículo en PDF publicado en dicha revista, a una biografía del autor en *Wikipedia* y a un recurso multimedia en *Vimeo* relacionado con el contenido del artículo. Un conglomerado de recursos relacionados semánticamente.

El uso del lenguaje de marcado RDF genera una serie de oportunidades potenciales para todos estos servicios, que se elevan a la máxima potencia cuando su utilización sigue los parámetros y recomendaciones del llamado Linked Data.

Cada URI debe ser complementada con una descripción del recurso, nombre u objeto al que identifica. Esta descripción se debe realizar mediante RDF. Esto implica que la información no será válida sólo para el lector, sino también para la máquina.

Finalmente, en esta descripción del recurso en RDF, se deben incluir enlaces a otros recursos, de forma que el usuario pueda navegar entre ellos, favoreciendo el descubrimiento de información. Además, las relaciones establecidas estarán tipificadas, es decir, existirá información semántica en la relación.

Eso se consigue gracias a una estructura basada en tripletes (sujeto: recurso; predicado: propiedad o relación; objeto: valor de la propiedad o recurso con el que se establece la relación), a partir de la cual los recursos se entrelazan semánticamente.

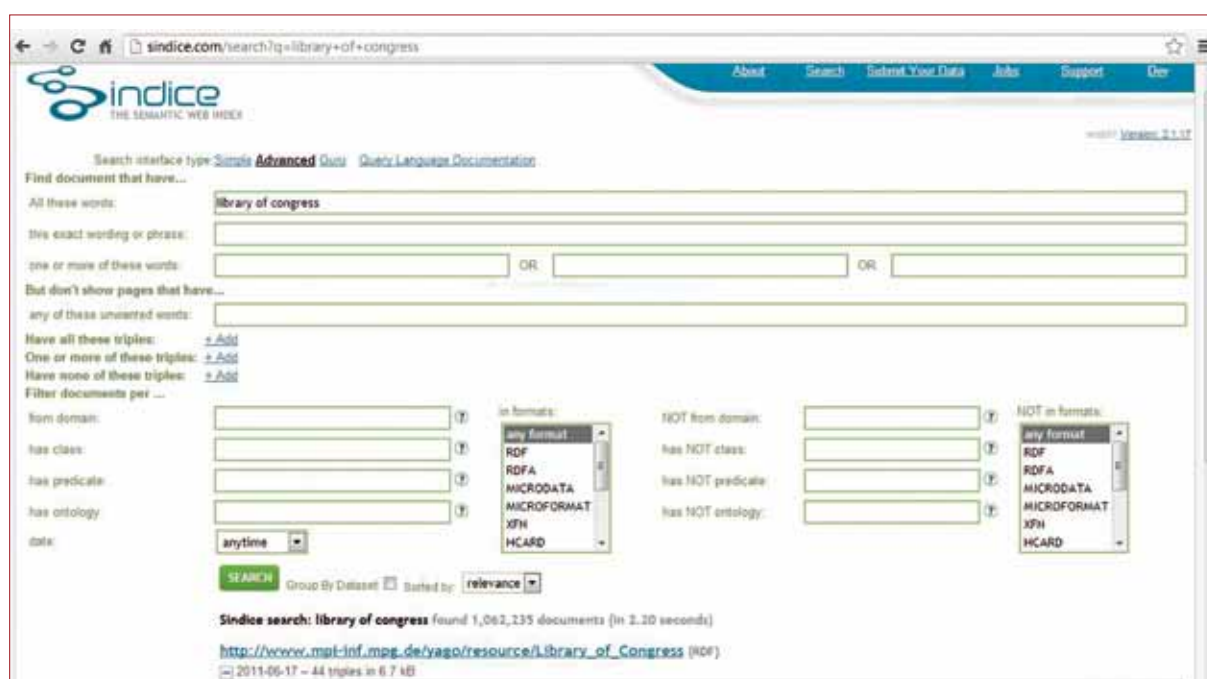
El verdadero poder de *Linked Data* proviene cuando una institución expone sus recursos siguiendo estas 4 pautas. Cuando una base de datos estructurada se transforma en RDF, explicitando relaciones entre recursos legibles por máquina. A esto se le conoce como *datasets*.

3.2. Algunos proyectos reales

Existen algunos *datasets* conocidos, como *DBPedia*³ (base de datos en RDF a partir de la información estructurada de los cuadros informativos de *Wikipedia*), o *Geonames*⁴, que expone información de nombres geográficos, descritos en RDF.

A partir de los *datasets*, se han desarrollado una serie de herramientas para buscar información a través de ellos, usando el lenguaje de consulta SPARQL. Entre éstos destacan navegadores, como *Marbles*⁵ o *Sigma*⁶, y buscadores como *Swoogle*⁷ o *Sindice*⁸, que son capaces de buscar tripletes RDF.

Igualmente, comienzan a aparecer aplicaciones construidas a partir de la información estructurada en RDF como método de integración de contenido. Entre estos proyectos destaca la BBC, con servicios como *BBC Programmes*⁹, *BBC Music*¹⁰ o *BBC Wildlife finder*¹¹, servicios que integran información proveniente de otros servicios, y la relacionan e integran con su propia información.



Búsqueda y visualización de información en Sindice

4. Conclusiones

La diversidad de servicios y plataformas en el entorno web genera una serie de oportunidades y retos en las bibliotecas, tanto en la gestión interna de información, como en su visualización y transferencia a los usuarios. Cada unidad de información puede decidir qué modalidad de las vistas encaja mejor en su sistema.

El uso del lenguaje de marcado RDF, y la tecnología asociada a este, generan una serie de oportunidades potenciales para todos estos servicios, que se elevan a la máxima potencia cuando su utilización sigue los parámetros y recomendaciones del llamado Linked Data.

Si deseamos llevar las bibliotecas a rendir más allá de la información y recursos propios, en un contexto de crisis en el que la apertura y reutilización de recursos de información son necesarios para maximizar los presupuestos, las diversas posibilidades en la generación de nuevos servicios puede permitir a las bibliotecas participar en el *mainstream* de datos que supone *Linked Data*, integrando servicios y plataformas que mejoren la experiencia tanto interna como de los usuarios finales. ▴

Referencias

- **Bizer, Chris; Cyganiak, Richard; Heath, Tom.** *How to Publish Linked Data on the Web*. [Report]. Fecha de consulta: 15-04-2011.
<http://www4.wiwiw.fu-berlin.de/bizer/pub/LinkedDataTutorial>
- **Dempsey, L. N.** (2007). "The network reconfigures the library systems environment". En: *Lorcan Dempsey Weblog*. Fecha de consulta: 15-04-2011.
<http://orweblog.oclc.org/archives/001379.html>
- **García García, Alicia.** (2012) Datos abiertos enlazados Linked Open Data (LOD) en documentación científica. Valencia: Universidad Politécnica de Valencia.
<http://riunet.upv.es/handle/10251/18272>
- **Pastor-Sánchez, Juan-Antonio** (2013) "Ontologías de control de autoridades en el ámbito de los datos abiertos enlazados" *Anuario ThinkEPI*, v. 7. [en prensa].
<http://www.thinkepi.net/ontologias-control-autoridades-datos-abiertos-enlazados>
- **Sellés-Carot, A.; Serrano-Cobos, J.** (2011) "Del OPAC extendido a la biblioteca expandida: al fin una realidad". *El profesional de la información*, v. 20, n. 4, pp. 460-464.
- **Troncy, Raphaël; Fialho, André, Hardman, Lynda; Saatho, Carsten** (2010). *Experiencing Events through User-Generated Media*.
http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-665/TroncyEtAl_COLD2010.pdf
- **Yang, Sharon Q.; Wagner, Kurt** (2010) "Evaluating and comparing discovery tools: how close are we towards next generation catalog?". *Library Hi Tech*, Vol. 28, n. 4, pp.690-709.
<http://www.emeraldinsight.com/journals.htm?articleid=1896437&show=abstract>

Notas

1. <http://linkeddata.org>
2. <http://esw.w3.org/LinkedData>
3. <http://dbpedia.org>
4. <http://www.geonames.org>
5. <http://marbles.sourceforge.net>
6. <http://sig.ma>
7. <http://swoogle.umbc.edu>
8. <http://sindice.com>
9. <http://www.bbc.co.uk/programmes>
10. <http://www.bbc.co.uk/music>
11. <http://www.bbc.co.uk/wildlifefinder>

Ficha técnica

AUTORES: Orduña- Malea, Enrique ; Sellés Carot, Alicia ; Serrano-Cobos, Jorge.

FOTOGRAFÍAS: <http://virtuoso.openlinksw.com>

TÍTULO: Estrategias y oportunidades en la generación de *Linked Data* en las bibliotecas

RESUMEN: Ante la diversidad de aplicaciones, servicios y tecnologías que cada día más se van integrando en las bibliotecas, se describen las oportunidades que la integración de servicios y plataformas suponen para las bibliotecas, basado en la utilización de la tecnología *Linked Data*, capaz de realizar inferencias con datos bibliográficos y no bibliográficos, además de permitir cambiar el paradigma tradicional de búsqueda-resultados-facetas, con el objetivo de que cualquier biblioteca o servicio de información pueda aplicarlo sin necesidad de pertenecer a un macroproyecto ni realizar costosas inversiones.

MATERIAS: *Linked Data* / Nuevas Tecnologías / Bibliotecas.